# Evaluate Location Features for Continuous Authentication with Machine Learning Experiments

Rossana M. C. Andrade[1]*, Marcio A. S. Correia[1], Carlos A. B. Carvalho[12], and Pablo Ximenes[3]

[1] Group of Computer Networks, Software Engineering and Systems
Federal University of Ceara, Fortaleza, Brazil
{rossana,marcio}@ufc.br
[2] Federal University of Piaui, Teresina, Brazil
candrebc@ufpi.edu.br
[3] Information Technology Company of Ceara (ETICE)
State Government of Ceara, Fortaleza, Brazil
pablo@ximen.es

**Abstract**

Traditional authentication mechanisms take some time, especially, from mobile users that access your devices several times but for short periods. Continuous authentication appears as an alternative, requiring less active user interaction. These mechanisms aim at reducing the time spent by users with authentication as well as raising the security level related to identity verification. Existing solutions use biometric data as the input of Machine Learning algorithms that identify if a user is authorized. It is possible to use a single biometric feature or combine some features. Then, it is necessary to analyze the impact of the features and machine learning algorithms when designing an authentication mechanism. In this context, this paper describes a methodology to design and perform experiments in which their results can be used in a decision-making process. We apply this methodology to evaluate features of outdoor location obtained from GSM or GPS, using algorithms available in the WEKA environment. Based on efficacy and efficiency measures, our experiments, using three datasets, indicate better results when GPS data and the algorithm J48 are used for authentication.

## 1 Introduction

Authentication solutions should be designed to balance security and usability. Due to limitations imposed by the interface of mobile devices, the use of passwords cannot be the best authentication solution. Besides, the duration and frequency of user authentication on mobile devices must be considered in the design of authentication mechanisms, because the users handle with these devices more frequently but for short periods of time [1]. Then, the usability cannot be ignored because the risk of the authentication mechanism being disabled.

Continuous authentication (also known as transparent or implicit authentication) appears as an alternative for user authentication, requiring less active user interaction [2]. In continuous authentication mechanisms, the user's identity is continuously monitored based on physiological and behavioral biometrics extracted from natural use of a device. Some examples of biometric features are touchscreen manipulation, typing, gait, face, voice, location, active mobile app, accessed websites, typed text, audible noise, received/sent text messages and calls. The user

*Researcher Scholarship - Technological Development and Innovative Extension (DT 2), sponsored by CNPq, Brazil

profile is built from patterns recognized from a single biometric feature or combined features. The patterns recognition is broadly based on machine learning algorithms [3]. The security of the authentication mechanisms relies on their effectiveness that is influenced by the features and algorithms used.

In this context, a decision-making process is essential to propose authentication mechanisms, observing what algorithms and features produce better results. The statistical analysis based on efficacy and efficiency measures allows the comparison between the algorithms and features. On the other hand, existing solutions did not follow guidelines such as the proposed by Alpaydin [3] to design them, enabling the experiments' reproduction and comparison with between solutions. Then, we propose, in this paper, a methodology to design and perform experiments in order to evaluate what algorithms and features are best suitable to be used by continuous authentication mechanisms.

It is interesting to mention the possibility to use this type of mechanism also by IoT devices. As an example, a presence sensor can identify automatically who entered the environment and send commands to the lights and the air conditioner, configuring them in accordance with the user's preferences.

GSM (Global System for Mobile Communications) or GPS (Global Positioning System) data have been used as location feature to authenticate users [4, 5]. However, it was not possible to infer what feature is most suitable for authentication. Then, we applied our methodology to evaluate location features. We detail here the results of this evaluation that includes also the analysis of machine learning algorithms and the merge of features, considering, for example, a user's trajectory. We use WEKA [6], a publicly available machine learning environment for experimentation that achieves wide acceptance in the academic world. Finally, we also use Geolife [7] and MIT Reality [8], two open datasets which have location collected from a large number of real mobile devices users.

Following sections present the background (Section 2); related work (Section 3); proposed methodology and case study (Section 4); results and analysis (Section 5); and finally, conclusions and future work (Section 6).

## 2 Background

Continuous authentication mechanisms have be proposed in literature to identify mobiles users without their active participation [1]. Mobiles devices have multiple sensors whose data can be collected and analyzed, allowing the user identification. Usually, the mechanisms are based on machine learning algorithms in which an authorized user provides samples of his/her features that are used to build the user profile. This process can be called registration and is followed by the authentication process [9]. In the authentication process, the unknown user identity is verified to allow user access to the system.

Although there are unsupervised learning algorithms [3], this scenario fits with the use of supervised classification algorithms. Then, the algorithm learns the pattern of the authorized user, based on the features extracts from the training set. Next, this pattern is compared with the features are extracted during the use of the device, allowing or not access. It is essential to perform experiments to analyze the effectiveness of the authentication. For this, a test set, containing features of authorized and unauthorized users, is used for classification, and the obtained results are compared with the expected results.

In this scenario, the training set contains only data of the authorized user, requiring a one-class classification algorithm such as One-class SVM [10]. We performed some experiments, using this algorithm, and obtained results well below that achieved by related work. However, one-class algorithms are not used by related work so that the training set contains data of authorized and unauthorized users. In this research, two-class were used, and each entry of the training set is labeled with authorized or unauthorized, indicating if this entry is referring to the authorized user.

In our experiments (see Section 5), each entry of the test set is classified also as authorized or unauthorized. In order to statistically analyze the results, it is necessary to collect some measures, considering also the expected results. The basic measures is True Positive ($TP$), True Negative ($TN$), False Positive ($FP$) and False Negative ($FN$). The first two measures show the hits of the classifier, indicating, respectively, the number of times the authorized and unauthorized users have been identified rightly. On the other hand, the latest measures inform the mistake of the classifier.

In order to enable the reproducibility and the comparison with other solutions, it is essential to follow some guidelines, during experiments, as the described by Alpaydin [3]. His guidelines specify the needs to: i) define the objectives of the experiments; ii) select the response variables; iii) choice of factors and levels (i.e., the controllable elements of an experiment, such as the algorithms and their configurations); iv) design the experiments, detailing their variations in accordance with, for example, with the used dataset and algorithm; v) perform the experiments; and vi) analyze the results. Our methodology is based on this guidelines and detailed in Section 4.

The repetition of experiments, using the same factors is essential to statistically analyze so that the effects of uncontrollable factors are reduced. In machine learning, a dataset is divided in different ways to repeat each experiment, using a cross-validation technique. Besides, in order to compare algorithms, the same subsets is used, modifying only the algorithms. Following the recommendation of Alpaydin [3], we use the 10-Fold Cross-Validated Paired T-Test, performing ten experiments with each algorithm. The dataset is divided in ten equals parts, and one part is used as the test set in each experiment while the others parts make up the training set.

# 3 Related Work

Patel et al. [2] present a review of the recent literature with major work involving continuous authentication. Among these work, [4] and [5] use location features. We performed a systematic literature review to identify other proposals that use the location for authentication. the Table 1 summarizes our findings, showing for each related work: i) the technology used to extract the users' location; ii) the extracted features; iii) the machine learning algorithm; iv) the evaluation measures; and v) details of the dataset.

Although we highlight only features related to user location, some solutions use other features to perform the user authentication, such as web browsing [5] and app usage [11]. However, the conducted evaluations do not describe the effect the impact of each feature in the efficiency and efficacy of the proposed mechanisms. Each mechanism focus in a single machine learning algorithm, such as K-NN (K-Nearest Neighbor) [13] or SVM (Support Vector Machine) [5], so that it is not evaluated what is the most suitable algorithm for continuous authentication, considering the selected features.

Besides, it is not possible to compare the solutions because a proper methodology was not applied for evaluation of the mechanisms. The absence of this methodology makes impossible the reproduction of the experiments that were performed not using open datasets. Then, although the related work presents good results under the observed scenarios, it is hard to conclude about the overall effectiveness of the proposed mechanisms

In this context, the related work does not provide elements that help in the decision-making

Table 1: Main analysed characteristics in related work.

| Paper | Tech. | Features | Learning | Measures | Datasets |
|---|---|---|---|---|---|
| Shi et al. [4] | GSM | CellID Sequences | Levenshtein Distance | TPR, FPR, Precision, and Recall | Collected by authors (7 users) |
| Fridman et al. [5] | GPS | Latitude and Longitude | SVM with Kernel RBF | TPR, FPR and ROC curve | Collected by authors (200 users) |
| Ramakrishnan et al. [11] | GPS | Latitude and Longitude | K-NN | TPR and FPR | Collected by authors (4 users) |
| Tang et al. [12] | GPS | Latitude, Longitude, and Time | Rule Based Classifier | TPR and FPR | Collected by authors (10 users) |
| Hayashi et al. [13] | GPS | Latitude and Longitude | K-NN | Others | Collected by authors (32 users) |

process used during the design of authentication mechanisms. It is possible to ask what algorithm and feature (e.g., GPS-based or GSM-based) are most appropriate. Another possible analysis is the effect of a set of features in the efficiency and efficacy of authentication. In this research, for example, we observe the impact of the time and collection of locations (trajectory).

# 4 Evaluation Methodology and Study Case

In this section, we describe the proposed methodology that is based on the guidelines of machine learning experiments [3]. We extend these guidelines, proposing two new steps, as presented in Figure 1. By exposing each step of our methodology, we also detail its execution in accordance with our study case, focusing on location features.

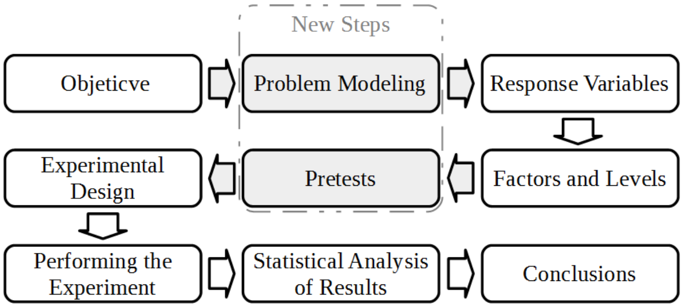


Figure 1: Methodology evaluation.

## 4.1 Objective

In machine learning, the definition of objectives is essential to plan the evaluation so that the executed experiments allow to achieve them. In this research, the objective of the experiments is compare the results obtained with outdoor location features, considering: i) GPS and GSM-based features; ii) machine learning algorithms; iii) effect of the use of time and trajectory in authentication; and iv) impact of the existence of users with similar behavior in the dataset.

### 4.2 Problem Modeling

In this step, the original problem (in our case, authentication) is mapped in a machine learning algorithm. This mapping is not in the scope of the guidelines [3] and helps in the choices of the type of algorithms to be used during the experiments. An authentication mechanism verifies if a user is authorized or not. As described in Section 2, our mapping results in datasets that contain samples of authorized and unauthorized users. Then, a two-class classifier can identify if some sample is authorized or not.

### 4.3 Response Variables

In this step, quality measures are selected to be used during the evaluation. The effectiveness is an important aspect to be analyzed, showing not only the success in the users' identification but also the security of a mechanism, blocking the unauthorized access. There are several measures (e.g., accuracy, precision and recall) to calculate the effectiveness of a classification [3].

Due to the space limitation, we present here only the accuracy that is computed as indicated in Equation 1, where the $N$ is the amount of entries of the test set, and $TP$ and $TN$ are, respectively, the number of true and false positives. This measure has been usually cited in the literature and focus on the hits of the classifier. However, we computed the other measures, obtaining also good results. Besides, efficiency measures must also be computed, showing the cost of computing resources such as CPU and memory. The energy consumption was not analyzed because this measure is not supported by WEKA.

$$Accuracy = (TP + TN)/N \tag{1}$$

### 4.4 Factors and Levels

The factors are the controllable elements that affect the results of the experiments. The main factors are the features, datasets and machine learning algorithms. The features are defined based on the objectives that include the analysis of the location, time and trajectory. Then, we specified four features sets ($FS$). The *FS#1* contains entries with one single location while the time of day that the location was collected is also indicated in the *FS#2*. Each entry of the *FS#3* represents one trajectory, composed by six consecutive locations, as recomended by Shi et al. [4]. The time of the last location was included in the *FS#4*.

Two open datasets were used to enable the analysis of the GPS and GSM data. While Geolife [7] contains users' data collected by GPS, MIT Reality [8] consists of data collected by GSM. We also built a dataset containing GPS data of two users with a similar behavior to evaluate the impact of this similarity.

The algorithms are selected based on the modeled problem and their availability in WEKA. In this research, we use the following classifiers: i) Decision Tree C4.5 (J48); ii) Support Vector Machine C-SVC (LibSVM); iii) Naive Bayes (NaiveBayes); and iv) 0-R (ZeroR). The levels of one factor are related to, for example, the different configurations of one algorithm. The pretests can help in the definition of the configurations to be used, reducing the size of the experiments without lose the quality of the results. In our experiments, the only change made in the standard setting was in LibSVM. For this algorithm, it was necessary to activate the normalization of the input features (-Z) and deactivate the use of the heuristics (-H).

### 4.5 Pretests

Besides helping in the definition of configuration parameters, the pretests are included to filter the datasets. The repeated information reduce the efficiency and does not contribute to the learning process, being necessary to remove redundant samples. In our scenario, it is necessary to discretize the space and time. As results of this step, the location scale was fixed in approximately 100 square meters and the time scale was fixed in one minute. Then, only the change of region is considered in the extraction of the trajectory. Besides, in order to obtain more realistic results, it is necessary to have some similarity between the users' behavior. So, only users of the same region (e.g., city) are analyzed together. Last, we remove users with less of 500 samples because the low quantity of samples is one of the causes of the underfitting [3].

### 4.6 Experimental Design

Based on the results of the previous steps, the experiments are planned, defining what combination of factors will be used. Although we have fixed the parameters of the algorithms, all possible scenarios are considered in our experiments, following the factorial design [3]. Then, we have 48 evaluation scenarios, making all combinations of the three datasets, four features sets and four algorithms. The statistic analysis depends on the repetition of experiments, and we generated 610 different experiments, using the 10-Fold Cross-Validated Paired T-Test.

### 4.7 Performing the Experiment

The equipment used for the experiments was a Lenovo Desktop ThinkCenter M91p with an Intel Core i5-2400 processor, a 8GB RAM and Microsoft Windows 7 Professional 64 bits SP1 operational system. The WEKA environment was chosen for the execution of the experiments conducted in this work. Python scripts were created to generate the ARFF file (used by WEKA) of each experiment. The files are available at https://goo.gl/DMtHn8 to allow the reproducibility of the experiments, filtering the original datasets and producing the training and test sets. This link contains also the tables with all measures of the statistical analysis.

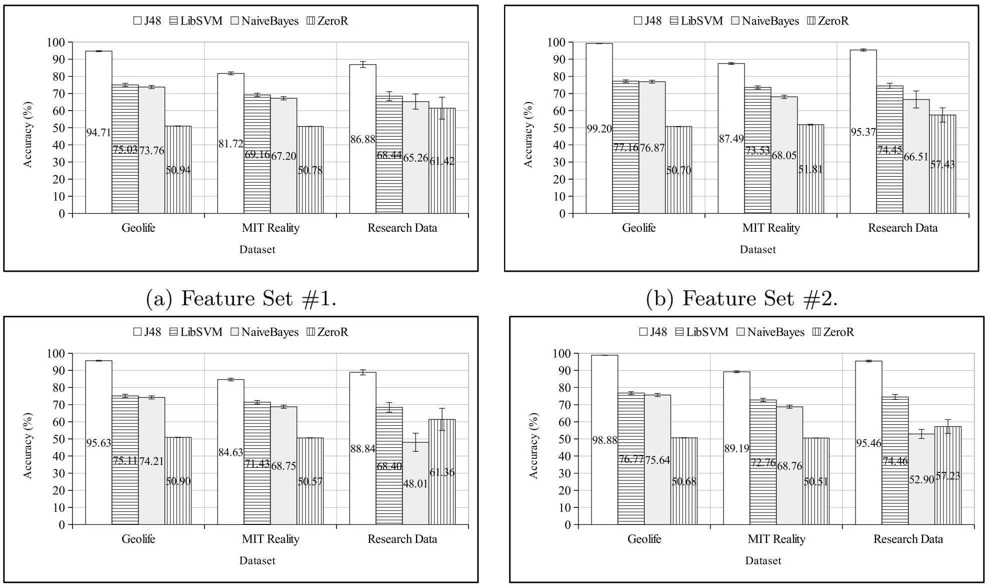
## 5 Statistical Analysis of Results

In this section, the results of this work will be presented. The results of the evaluations form the classifiers are presented in a graph, with a confidence interval of 95%. The comparisons between the obtained results by the classifiers considered a significance level of 0.05.

Figure 2 presents only the accuracy of the classifiers, grouped by dataset. Comparing the results obtained by the classifiers, it is possible to observe that the J48 reached the greatest accuracy among the algorithms, followed by LibSVM that had a significantly lower accuracy.

We also found evidence that the use of GPS data produces better results when compare with GSM data, extracted from dataset MIT Reality. Although the datasets are different, it is important to note that the behavior of the users in our GPS-based dataset (Research Data) is similar and, even so, the obtained results were better than the results of the MIT Reality.

The effect of the use of time and trajectory in authentication is another aspect to be analyzed. It is possible to view, in Figures 2a and 2b, the improvement of the results when the time and location are used as features. However, observing also Figure 2d, the inclusion of the trajectory generated best results, considering GSM data, and worse results using GPS data. On the other hand, there is a reduction in efficiency as shown in Figure 3.



(a) Feature Set #1.

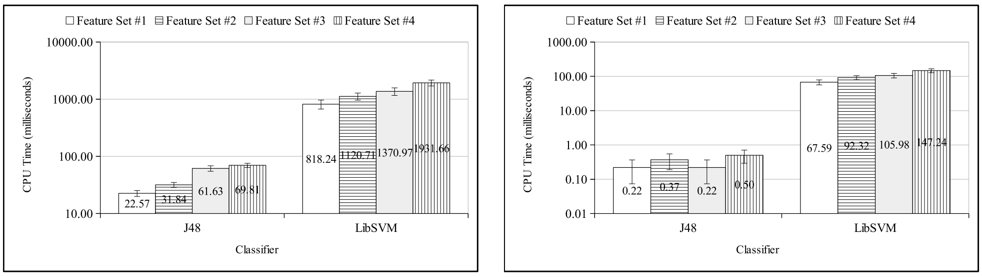
(b) Feature Set #2.

(c) Feature Set #3.

(d) Feature Set #4.

Figure 2: Accuracy results

The efficiency is other an important aspect, especially considering that the authentication is a real-time procedure. Due to the restriction of space, Figure 3 display only results obtained by the J48 and LibSVM classifiers, in accordance with the results obtained from the Geolife dataset. It is possible to highlight the good results obtained with the J48.



(a) CPU Time for training.

(b) CPU Time for test.

Figure 3: Performance results

# 6 Conclusions and Future Work

In this work, we propose a methodology to evaluate the use of machine learning algorithms by continuous authentication mechanisms. We applied our methodology to compare algorithms and different strategies of outdoor location that can be used by these mechanisms. The evidence obtained from the experiments indicated that Decision Tree C4.5 (J48) is the machine learning algorithm that offers the best efficacy and efficiency among the classifiers evaluated. Besides, analyzing the results by location techniques, the evidence indicated that the location features retrieved from the GPS data offer better results. Finally, considering the evaluated features sets, the evidence indicated that the combination of location features related to time and trace of last locations degrades the accuracy with GPS and improves the accuracy with GSM.

Opportunities for future work include the performing of experiments using different configuration of the evaluated algorithms, other algorithms (e.g., K-NN), other biometric features (e.g., face) and the development of an environment for conducting experiments in mobile devices. The proposed methodology can also help to define the most appropriate techniques and features in the design of new authentication mechanisms, including the solutions for IoT environment.

# References

[1] Heather Crawford. Adventures in authentication–position paper. In *Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, 2014.

[2] Vishal M Patel, Rama Chellappa, Deepak Chandra, and Brandon Barbello. Continuous user authentication on mobile devices: Recent progress and remaining challenges. *IEEE Signal Processing Magazine*, 33(4):49–61, 2016.

[3] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2014.

[4] Weidong Shi, Jun Yang, Yifei Jiang, Feng Yang, and Yingen Xiong. Senguard: Passive user identification on smartphones using multiple sensors. In *2011 IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 141–148. IEEE, 2011.

[5] Lex Fridman, Steven Weber, Rachel Greenstadt, and Moshe Kam. Active authentication on mobile devices via stylometry, application usage, web browsing, and gps location. *IEEE Systems Journal*, PP(99):1–9, 2015.

[6] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[7] Yu Zheng, Xing Xie, and Wei-Ying Ma. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 33(2):32–39, 2010.

[8] Nathan Eagle and Alex Sandy Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.

[9] Nathan Clarke. *Transparent user authentication: biometrics, RFID and behavioural profiling*. Springer Science & Business Media, 2011.

[10] Larry M Manevitz and Malik Yousef. One-class svms for document classification. *Journal of Machine Learning Research*, 2(Dec):139–154, 2001.

[11] Arun Ramakrishnan, Jochen Tombal, Davy Preuveneers, and Yolande Berbers. Prism: Policy-driven risk-based implicit locking for improving the security of mobile end-user devices. In *Proceedings of the 13th International Conference on Advances in Mobile Computing and Multimedia*, pages 365–374. ACM, 2015.

[12] Yujin Tang, Nakazato Hidenori, and Yoshiyori Urano. User authentication on smart phones using a data mining method. In *Information Society (i-Society), 2010 International Conference on*, pages 173–178. IEEE, 2010.

[13] Eiji Hayashi, Sauvik Das, Shahriyar Amini, Jason Hong, and Ian Oakley. Casa: context-aware scalable authentication. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, page 3. ACM, 2013.